Research Article

Psychological Science
22(1) 71–79

DOI: 10.1177/0956797610391098
http://pss.sagepub.com

Round Numbers as Goals: Evidence From Baseball, SAT Takers, and the Lab

Devin Pope[1] and Uri Simonsohn[2]
[1]The Booth School of Business, University of Chicago, and [2]The Wharton School, University of Pennsylvania

Abstract

Where do people's reference points come from? We conjectured that round numbers in performance scales act as reference points and that individuals exert effort to perform just above rather than just below such numbers. In Study 1, we found that professional baseball players modify their behavior as the season is about to end, seeking to finish with a batting average just above rather than below .300. In Study 2, we found that high school students are more likely to retake the SAT after obtaining a score just below rather than above a round number. In Study 3, we conducted an experiment employing hypothetical scenarios and found that participants reported a greater desire to exert more effort when their performance was just short of rather than just above a round number.

Keywords

judgment, decision making, reference points, goals

Received 5/7/10; Revision accepted 6/25/10

Where do people's reference points come from? Previous research shows that goals (Heath, Larrick, & Wu, 1999; Larrick, Heath, & Wu, 2009), expectations (Feather, 1969; Mellers, Schwartz, Ho, & Ritov, 1997), and counterfactuals (Kahneman & Miller, 1986; Medvec, Gilovich, & Madey, 1995; Medvec & Savitsky, 1997), among other possibilities, can all act as reference points when people evaluate outcomes. We hypothesized that because round numbers are "cognitive reference points" in numerical scales (Rosch, 1975), they also act as reference points in the process of subjectively judging outcomes. We conjectured that, as a result of this process, people with a performance score just short of a round number will be more likely to exert effort to improve their measured performance than people with a performance score just above a round number will be.

We tested this prediction by studying professional baseball players and students taking the SAT, and the results showed the expected relationship between performance relative to a round number and exertion of effort to increase measured performance. The effects we documented were significant, both statistically and in practical terms. For example, we found that professional batters are nearly 4 times as likely to end the season with a .300 batting average as they are to end the season with a .299 average. Similarly, we found that high school juniors were at least 10 to 20 percentage points more likely to retake the SAT if their total score ended in 90 (e.g., 1190) than if it ended in the most proximate 00 (e.g., 1200). We obtained consistent results in an experiment with hypothetical scenarios that ruled out some alternative explanations for the field studies.

In examining the role of goals as reference points, the research presented here is most closely related to the studies of Medvec and Savitsky (1997) and Heath et al. (1999). Medvec and Savitsky showed that participants report greater satisfaction when they imagine barely meeting a category of performance than when they imagine comfortably exceeding it. Heath et al. argued that explicitly set goals act as reference points and that, among other consequences, this use of goals as reference points leads people to express greater motivation for improvement when they are just short of meeting a goal. We contribute to these findings by demonstrating that a round number—a goal that has not been explicitly set and that is not attached to a direct consequence—is a powerful motivator of behavior, both inside and outside the laboratory.

Corresponding Author:
Uri Simonsohn, The Wharton School, University of Pennsylvania, 3730 Walnut St., 500 Huntsman Hall, Philadelphia, PA 19104
E-mail: uws@wharton.upenn.edu

Study 1: Baseball Players

Method

In Study 1, we examined the behavior of professional baseball players. Sports performance data have the general advantage of being available and of interest to players themselves, and baseball has an additional advantage of possessing a particularly salient measure of performance that varies with great granularity: batting average.[1]

We examined how players respond to their season's batting average being just below versus just above a round number. For ease of exposition and analysis, we refer to performance that is exactly equal to a round number as being above that number. We tested two predictions arising from the hypothesis that round numbers act as motivating goals: (a) that fewer than the expected number of players will end the season with a batting average just below a round number, and more than the expected number of players will end the season with a batting average above it, and (b) that players with batting averages that are very close to a round number as the season is ending will adjust their behavior to attempt to end the season with an average above the round number.

To test these predictions, we obtained play-by-play data for all Major League Baseball players from 1975 to 2008.[2] To ensure a sufficiently granular batting average, we restricted our sample to players who had at least 200 at bats during the season (reducing our overall sample from 11,430 player-seasons to 8,817 player-seasons). As batting averages for professional baseball players almost never drop below .200 or go above .400, we focused on batting averages around .300, the roundest number in that range.

Results

Most of our statistical analyses employed the full range of batting averages in the data. However, for ease of exposition and analysis, we graphically present our results in the range from .280 to .320 (n = 3,083 season × player observations). Figure 1 depicts the relative frequency of specific batting averages at the end of the baseball season and with five plate appearances left in the season. We included the latter distribution as a control to account for a possible mechanical blip in the frequency of .300 arising from rounding to three decimals, and for the possibility that the distribution of batting ability is discontinuous at .300.

Our findings were consistent with the notion that players use round numbers as goals. Season averages were markedly less likely to be just below .300 than just above .300. For example, the percentage of players ending the season with a .298 or .299 (0.97%) was lower than the percentage of players ending with a .300 or .301 (2.30%), $Z = 7.35$, $p < .001$. Furthermore, the marked increase between .299 (0.38%) and .300 (1.40%), $Z = 3.54$, $p = .001$, which implies that batters are nearly 4 times as likely to end with a .300 than with a .299 average, is the only increase between consecutive observations in Figure 1 with a p value lower than .10. No such pattern was observed for the control distribution of season average with five plate appearances left in the season. These results suggest that players find a way in their last few scheduled plate appearances to ensure that they finish above .300. We assessed how they do this by testing our second prediction (see Fig. 2).[3]

Figure 2a shows that in the last scheduled plate appearance of the season, a higher percentage of players achieved a base hit when their batting average was .298 or .299 (35.2%) than when their batting average was .300 or .301 (22.4%), $Z = 2.36$, $p = .018$. Furthermore, the percentage of players with a batting average of .299 who got a base hit (43%) was higher than the overall percentage of all other players represented in the figure who did so (22.8%), $Z = 3.62$, $p < .001$.[4] This provides further evidence that the high frequency of .300 averages in Figure 1 arose from actions batters took at the end of the season, rather than from mere rounding or some other mechanical cause.

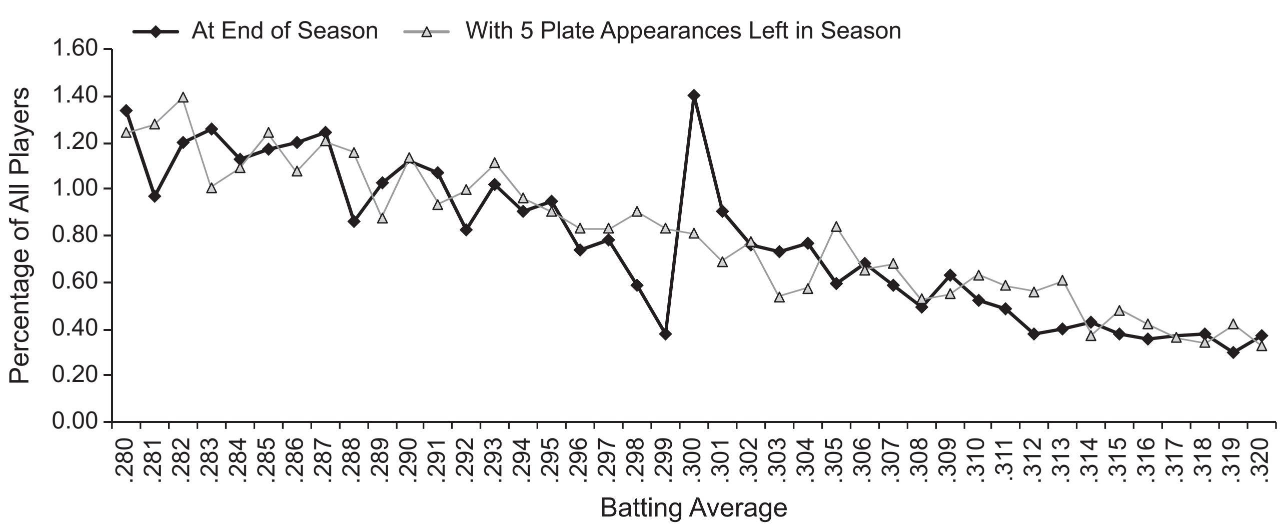

Fig. 1. Relative frequency of batting averages among Major League Baseball players between 1975 and 2008. Batting averages at the end of the baseball season and with five plate appearances left in the season are shown. The graph includes only player-seasons with at least 200 at bats.

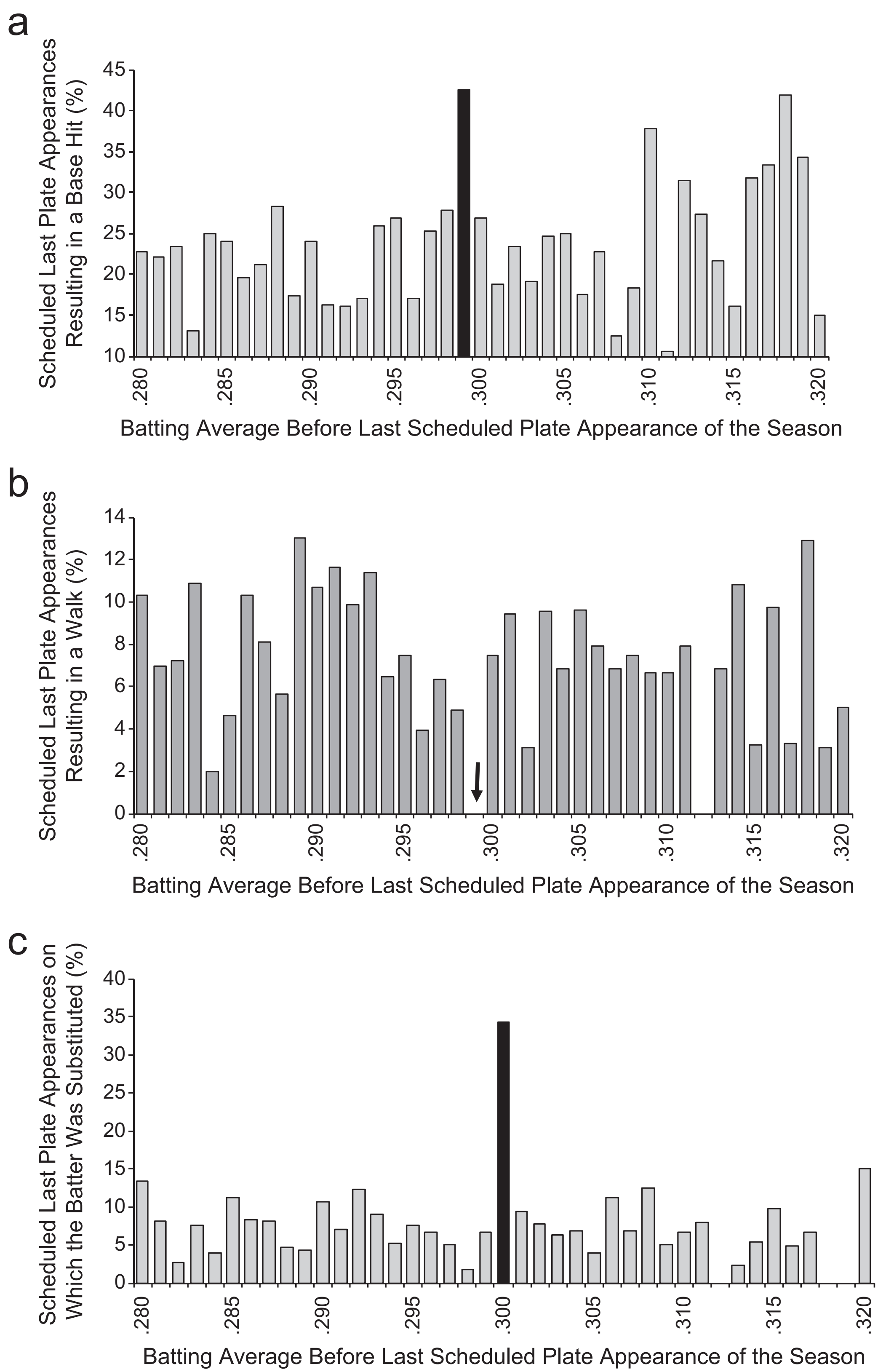

Fig. 2. Outcome of the last scheduled plate appearance of the season: percentage of plays resulting in (a) base hits, (b) walks (which cannot increase batting average), and (c) batter substitutions (pinch hitter brought in). Bars involved in tests of predictions are highlighted in black. The arrow in (b) emphasizes that not a single player with a batting average of .299 walked.

There are several specific actions players could have taken to achieve the results depicted in Figures 1 and 2a. First, players whose averages were just below .300 might have chosen to walk less because walks cannot increase a batting average. Figure 2b shows that players were in fact less likely to walk when their batting average was .298 or .299 (2.5%) than when their batting average was .300 or .301 (8.6%), $Z = 2.14$, $p = .032$. Furthermore, not a single player (out of 61) walked when his season average was .299.

Second, if players' averages were just above .300, they might have ended their season a few plate appearances earlier. One way to achieve this would be by having a substitute (pinch hitter) bat instead of them. Figure 2c shows that players with a .298 or .299 average were less likely to be replaced by a pinch hitter on their last scheduled plate appearance (4.1%) than were those with an average of .300 or .301 (19.7%), $Z = 3.85$. $p < .001$. Furthermore, the rate of substitution for players with an average of .300 (34.3%) was by far the highest of all players included in the figure, whose overall percentage of substitution was 7.0%, Z = 8.29, $p < .001$. Similarly, players might have skipped the entire last game or last few games if their batting average was already at .300. The high base-hit percentage for .299 batters depicted in Figure 2a, therefore, is likely to be at least partially explained by batters ending their season prematurely when they obtained a desired goal, rather than by players actually batting more successfully when they had not.

Discussion

Overall, the behavior of baseball players proved consistent with the hypothesis that a round number, such as a batting average of .300, can act as a goal that influences behavior. Our analyses of the baseball data have two notable limitations: There was only one relevant round number, and players' actions to improve performance took place on the last plate appearances of the seasons and hence had relatively minor consequences. In our next study, we aimed to address both of these limitations.

Study 2: Retaking the SAT

Method

The SAT is a standardized test for college admission in the United States. Until 2006 (and for the entirety of our sample period), SATs were scored between 400 and 1600, in intervals of 10. Students are allowed to retake the test, and a large percentage of them do (about 50%, according to Vigdor & Clotfelter, 2003). We conjectured that if round numbers act as performance goals for test takers, then individuals scoring just below a round number would be more likely to retake the test than those scoring just above a round number would.

In Study 2, we used data from the College Board's Test Takers Database, a restricted-use data set one of the authors obtained for an earlier research project (Pope & Pope, 2009). Our data set was a random sample of 25% of all SAT test takers graduating between 1994 and 2001. It also included, for those same years, 100% of all SAT test takers from California and Texas, and all test takers self-reported as African American or Hispanic ($N = 4,323,906$).

The data set included only the score and date of the final test each student had taken; it did not include information on whether the student had taken the test before. Hence, we did not directly observe which students retook the SAT. Instead, as we describe in detail later in this section, our analyses identified retaking rates by looking at the SAT-score distributions for juniors and seniors. Juniors and seniors account for 99.5% of the data, so we focused on these two groups only.

Results and discussion

Test retaking. Figure 3 shows the distribution of SAT scores separately for juniors and seniors. The distribution for juniors is centered to the right of the distribution for seniors, indicating that, on average, the juniors (who did not retake the SAT as seniors) did better on the SAT than the seniors did. In this study, we were primarily interested in gaps in the frequency of scores around round numbers for juniors (as compared with seniors), from which we inferred the rate of test retaking.

The majority of high school seniors who take the SAT do not have the opportunity to receive their scores and then retake the test before they send out their college applications. As expected, we found that the distribution of seniors' scores was smooth. The majority of juniors who take the SAT, in contrast, have the opportunity to see their scores and then have the option of retaking the test before sending out college applications. Thus, if students are more likely to retake the exam if they score just below rather than just above a round number, we would expect to see discrete jumps in the frequency of juniors with scores below and above round numbers. Indeed, such a pattern can be seen in Figure 3.

The most visually striking gap in the data occurred between the scores of 990 and 1000. There were 18,134 juniors who obtained a score of 990 and 20,057 juniors who obtained a score of 1000, a difference of 1,923 students, or 10.6%. Among seniors, the corresponding numbers were 58,714 and 58,716, respectively; that is, just 2 more seniors obtained a score of 1000 than a score of 990. In other words, there are roughly 11% too few 990 scores among juniors when one uses the seniors' data as a control. There is a problem with this intuitive comparison, however, as the distribution for seniors' scores peaked before that of juniors: We address this problem in subsequent statistical analyses by employing a more conservative baseline for expected score frequencies.

Although these gaps in the distribution of juniors' scores are apparent to the naked eye for most round numbers above 700 in Figure 3, the gaps appear to be smaller for scores that are further from 1000. These visual comparisons are misleading, however, because their denominator also gets smaller as the numbers move away from 1000.

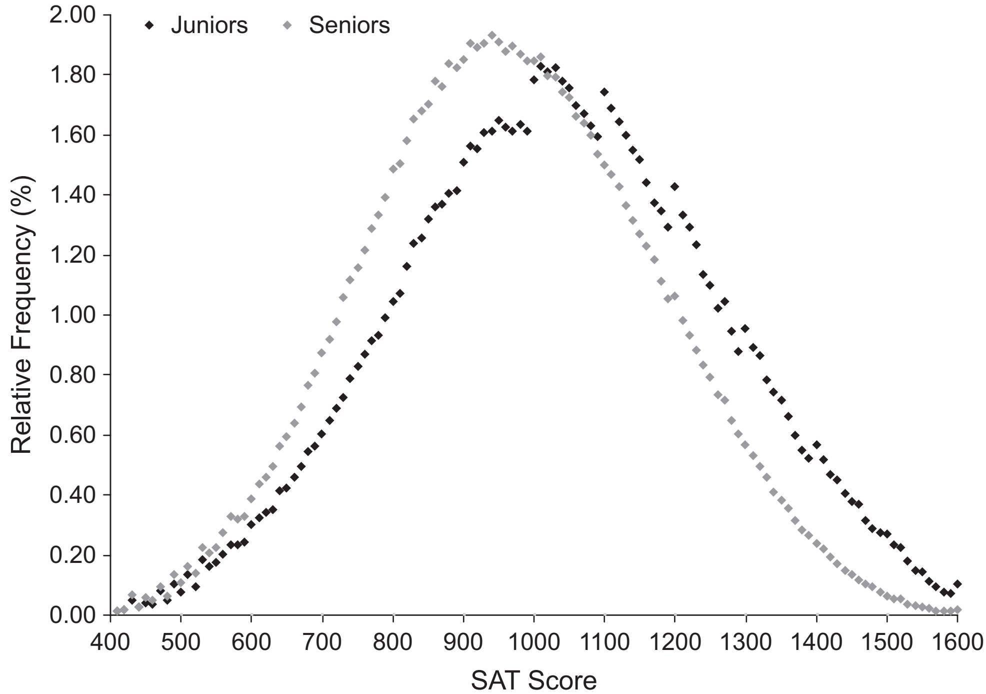

Fig. 3. Distribution of SAT scores of high school juniors and seniors. Only the final SAT score for each student was included in this data set.

To account for this, we plotted the ratio of relative frequencies (see Fig. 4) between consecutive scores (the slope of the density function). For example, Figure 4 shows that the ratio of relative frequencies at a score of 1000 is 1.11 for juniors but 1.00 for seniors. The figure shows sizable gaps for juniors at every round number between 900 and 1500, with the gaps between 1000 and 1400 being particularly large. From 1390 to 1400, for example, the ratio of relative frequencies is 1.08 for juniors but 0.89 for seniors, implying at least a roughly 20 percentage points greater retaking rate among juniors scoring 1390 than juniors scoring 1400.

As already mentioned, our first comparisons suffered from the problem that the distribution for seniors peaked before that for juniors. A simple way around this problem is to exploit the fact that for scores above 1000, the frequency of juniors who get a given score should drop as the score in question increases (e.g., there should be fewer juniors with a score of 1250 than with a score of 1240).

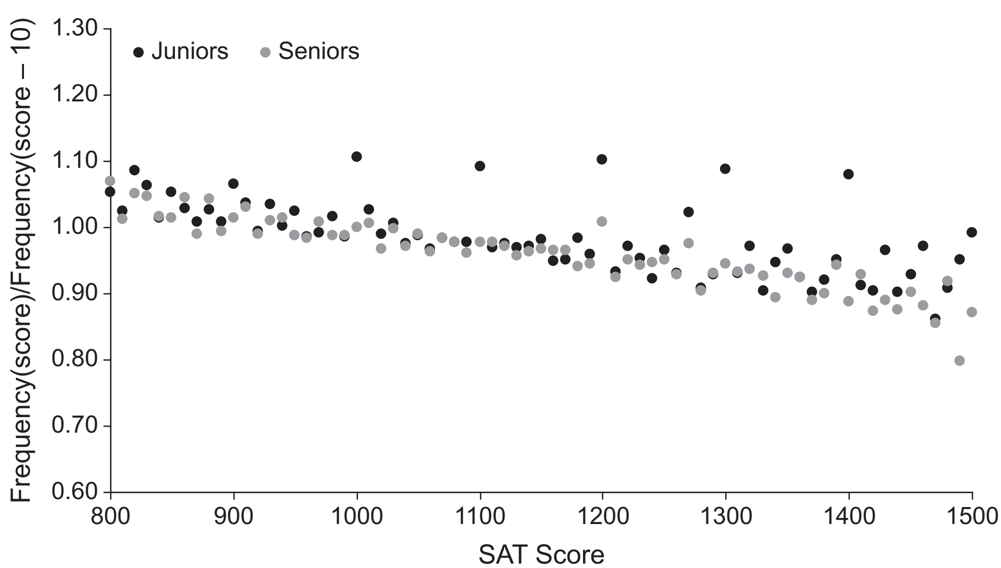

Fig. 4. Ratios of relative frequencies between consecutive SAT scores, for high school juniors and seniors.

We therefore conducted a conservative test of the prediction that scoring just short of a round number increases the odds of retaking the SAT, by testing the null hypothesis that the relative frequency of students who obtain a score that is a round number is the same as the relative frequency of students who obtain the immediately lower score. We conducted simple difference-of-proportions tests to assess whether we could reject the null hypothesis that these two proportions are equal. For example, there were 1,656 more juniors with a score of 1100 than with a score of 1090 (1.74% vs. 1.60%), $Z = 8.61$, $p < .001$, which we interpret as favoring the notion that falling short of 1100 provided greater motivation to retake the SAT than getting 1100 did. Analogous conservative calculations rejected the null hypothesis for scores of 1200, 1300, and 1400, all ps < .001.

In sum, we found a systematic pattern consistent with students being more likely to retake the SAT if their score is just short of a round number. We interpret this as evidence of SAT takers using round numbers as implicit goals for performance. A related (though alternative) account involves test takers, either correctly or incorrectly, believing that surpassing a round number disproportionately increases their odds of being admitted to college or of receiving financial aid. This is not implausible considering that (a) universities and scholarship programs often do impose minimum SAT thresholds for consideration (although probably not as high as 1300 or 1400) and (b) the admission process has a considerable component of subjective judgment that may respond to round numbers as well. Although this account is very closely related to the notion that SAT takers use round numbers as goals (it posits that SAT evaluators do), we attempted to distinguish between these two possibilities by conducting additional analyses of the SAT data and undertaking an experiment (Study 3) that ruled out such concerns by design.

Score sending. The data set that we used contained information about the colleges to which students sent their scores. We used this information to indirectly test whether the students believed that a score just short of a round number is disproportionately worse than a score just above a round number. Intuitively, if this is the case, students with scores just above a round number would send their scores to different schools than students with scores just below the same round number would. To measure the quality of schools to which scores were sent, we first computed the average SAT score sent to each school by all applicants in the sample. We then computed the mean of these school averages for the set of schools to which each applicant sent scores. In short, we used the average SAT score of other students who sent their scores to particular schools as a proxy for the quality of schools to which a given test taker chose to apply.[5]

Figure 5 shows the average quantity and quality of schools seniors' SAT scores were sent to, as a function of SAT score.[6] Both lines slope upward, indicating that students with higher SAT scores applied to more and better schools. More important, the figure shows that the quality and quantity of schools to which scores were sent increased smoothly as scores passed round numbers. We interpret this as suggesting that round-number scores influence retaking decisions because of their effect on motivation to improve measured performance, rather than because of their direct impact on the outcomes students believe they can obtain with such scores.

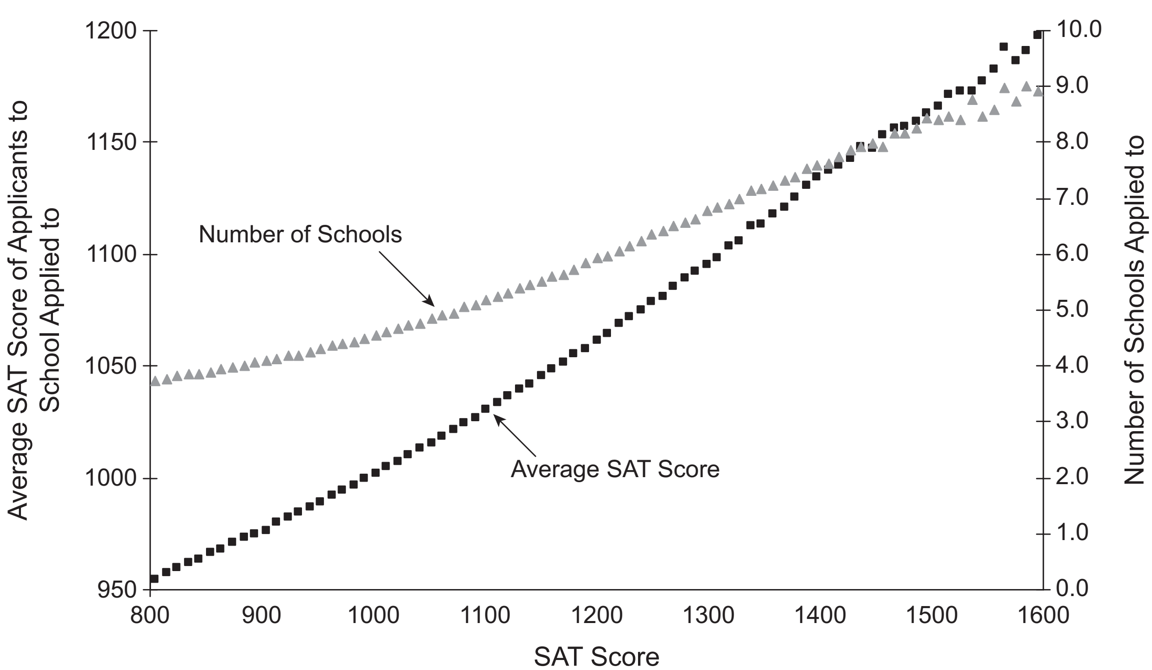

Fig. 5. Average quantity and quality of schools to which high school seniors sent their SAT scores as a function of the SAT score obtained.

Admission decisions. Figure 5 provides indirect evidence that test takers do not believe their odds of admission change discontinuously by passing a round score. Data on admission decisions are required in order to assess whether or not these beliefs are correct. Although there is no centralized data set of admission decisions across universities, one of the authors of this article has used admissions data from two institutions in previous (and completely unrelated) work, and we used those data in this study to partially address this question.

The first data set was from a highly competitive private university for which just over 1,100 undergraduate admission decisions were available (originally used in Simonsohn, 2007). We estimated regressions with the admission decision as the dependent variable, and with dummy variables for SAT scores as key predictors, controlling for other academic characteristics of the applicants. In other words, we obtained conditional average admission probabilities for different SAT scores. We tested the null hypothesis that the probability of admission changes by the same amount for a given increase in SAT score, whether the increase is from just below a round number or from just above that round number (e.g., that the probability of admission increases by the same amount between test scores of 1390 and 1400, as between test scores of 1400 and 1410). This is a conservative test, because we would expect higher scores to have smaller marginal effects on admission decisions.[7] We tested this null hypothesis for scores of around 1200, 1300, 1400, and 1500, and failed to reject the null hypothesis in all four cases (*n*s = 55, 97, 144, and 87, respectively; *p*s = .96, .99, .20, and .92, respectively).

The second data set we used was from a business school and hence consisted of Graduate Management Admission Test (GMAT) rather than SAT scores. We tested the equivalent null hypothesis for scores of around 600 and 700 (the maximum score on the GMAT is 800) and again failed to reject the null hypothesis (*n*s = 3,432 and 1,404, respectively; *p*s = .09 and .93, respectively).

In sum, the admissions data suggest that students' probability of admission does not disproportionately change as scores pass round numbers, and the score-sending data suggest that students behave accordingly. Our next and final study was a scenario experiment that further eliminated concerns that third parties might drive the motivating effects of round numbers.

Study 3: Scenarios

Method

This study was part of a sequence of unrelated experiments conducted at a behavioral laboratory. Participants were given a flat payment for their participation. The study was computer based and presented each participant with three scenarios, always in the same order. Each scenario presented participants (N = 172) with a situation involving a measure of their own performance. They were given feedback on their performance before the task had concluded and were asked how motivated they thought they would be to improve. The sole manipulation in Study 3 was the level of performance indicated. Six different versions of each scenario were created for this experiment, which had a 3 (distance from round number: far below, just below, just above) × 2 (round number: low, high) design. To ensure a between-subjects design, we assigned each participant to the same condition for all three scenarios (e.g., a participant would be given feedback indicating performance that was far below the low round number for all three scenarios).

The scenarios were as follows:

1. Imagine that in an attempt to get back in shape, you decide to start running laps at a local track. After running for about 30 minutes and having done (18/19/20/28/29/30) laps, you start feeling quite tired and are thinking that you might have had enough. How likely do you think it is that you would run one more lap?
2. Imagine you are participating in a basketball tournament, and that your current free-throw average is (48.2/49.2/50.2/58.2/59.2/60.2)%. Before your next game you calculate that if you made two free throws (and missed none) your season average would go up by just over 1%, to (49.3/50.3/51.3/59.3/60.3/61.3)%. During the game, you are fouled and walk to the line to shoot two free throws as your team is ahead by 5 points. How motivated do you think you would be to make those free throws?
3. Suppose that you got a temp job to make some extra money. The job is tedious as it consists of copying and pasting from Acrobat (.pdf) files into an Excel spreadsheet, and then manually fixing cells that did not copy properly. You get paid roughly $3.50 per table copied. You are considering whether to do one more table or whether to head home directly. You check out your computer screen and see that today you have so far made $(84.16/88.16/92.16/94.16/98.16/102.16). How likely do you think it is that you would do another table before going home?

Respondents answered these questions employing 9-point Likert scales. In Scenarios 1 and 3, the scales ranged from 1, *extremely unlikely*, to 9, *extremely likely*. In Scenario 2, the scale ranged from 1, *not at all motivated*, to 9, *extremely motivated*.

Results and discussion

Our dependent variable was the average score for each participant across the three scenarios (M = 6.34, SD = 1.45, minimum = 2.3, maximum = 9; see Fig. 6).

Factorial analysis of variance revealed a main effect for distance from the round number, $F(2, 171) = 15.31$, $p < .0001$; there was no main effect for whether the round number was high or low, $F(2, 171) = 0.20$, $p = .656$, and no statistically significant interaction between the two, $F(2, 171) = 1.02$, $p = .362$.[8] The main effect for distance from the round number arose because people reported greater motivation for improvement when their performance was just below a round number

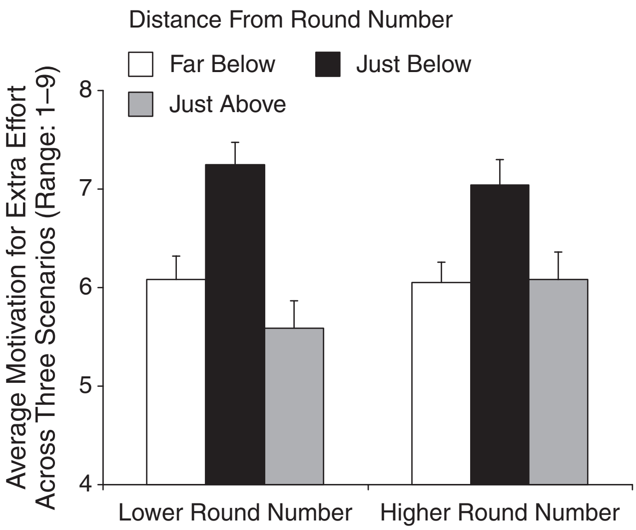

Fig. 6. Results of Study 3: average reported motivation to exert more effort as a function of current performance. Results were averaged across three scenarios in a 3 (distance from round number) × 2 (high or low round number) design. A given participant was assigned all three scenarios in the same condition. Error bars indicate 1 *SEM*.

(M = 7.14) than when their performance was far below a round number (M = 6.06), $t(112) = 4.61$, $p < .001$, or just above a round number (M = 5.84), $t(112) = 4.94$, $p < .001$. The fact that the round-number level did not influence motivation means that participants were not making inferences about overall effort or exhaustion based on that level, and this further suggests that they responded to distance from a round number because of its motivating effect rather than because of other inferences.

Conclusion

The studies presented in this article show that round numbers in performance scales act as goals that motivate individuals whose current measure of performance is just below a round number to improve their measured performance. These findings contribute to the literature on reference points in at least two ways. First, they suggest a new source of reference points that are naturally occurring and present in many different situations. Second, they provide evidence, which is often difficult to obtain, that reference points matter for real-life decisions.

We believe that the fact that round numbers can act as goals leads to at least two interesting questions for future research. The first is how these implicit goals, round numbers, interact with explicitly set goals. It is common for explicit goals to be set at round numbers; if round numbers are goals even without incentives, could it be that explicit goals set at other numbers would lead to greater overall motivation? For instance, if people would naturally want to lose 30 pounds, would it be more effective for them to set an explicit goal of losing 33 pounds rather than 30 pounds? Or is it the case that explicit goals are particularly effective if they are combined with implicit ones? The second question for future research concerns the precise mechanism by which round numbers are motivating. One possibility is that people performing below a round number subjectively assess the odds of better performance in the future as greater than do people performing at just above a round number (a judgment-mediated effect). Another possibility is that performing short of a round number is simply disproportionately aversive (a utility-mediated effect). It seems likely that both mechanisms play a role.

Acknowledgments

Authorship order was determined randomly. The authors would like to thank Dean Foster, Jack Hershey, Shane Jensen, Alan Schwarz, Abraham Wyner, and seminar participants at The Wharton Sport Business Initiative Conference in Philadelphia (November 2009), Wharton's Decision Processes Seminar in Philadelphia (June 2010), and the Behavioral Decision Research and Management Conference in Pittsburgh (June 2010) for their useful comments and suggestions.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. Batting average is the number of times a player successfully hits the ball divided by the number of at bats. The Baseball Almanac (n.d.) refers to it as "easily the most common statistic in baseball and the most understood."
2. Play-by-play data correspond to Retrosheet's (n.d.) Regular Season Event Files for 1975 through 2008. Season-level results were computed by aggregating the individual plays. Our calculations were validated against season-level data from a baseball archive (Lahman, n.d.).
3. Most players are likely to know that their last plate appearance is in fact their last. In our data, for instance, there was a 79% chance that a player who approached the plate in the eighth or ninth inning did so for the last time in that game. If players are uncertain about which plate appearance is their last, this would bias our results toward a zero effect.
4. Note that our calculations of the outcomes of the last scheduled plate appearance of the season included in the denominator all possible outcomes: outs, hits, walks, and substitutions. For example, Figure 2a shows that 27% of batters with a .300 average obtained a base hit on their last scheduled plate appearance. This is different from the batting average of the last at bat, which excludes both walks and substitutions from the denominator. We chose to include these to avoid the selection bias that would be introduced by batters who decided whether to bat, or walk, as a function of their current batting average. This correction, however, does not take into account further selection bias arising from batters who may decide to miss entire games after obtaining their desired batting average.
5. This measure correlated highly with that provided by the rank of top-50 schools ("America's Best Colleges," 1997) for 1998 (the midyear of our sample), $r = .71$, $p < .0001$.
6. We analyzed score-sending behavior by seniors because the sample of juniors had selection bias; for juniors, we observed school choices only by those students who felt their score was high enough that they did not need to retake the test.

7. Because of the small sample size, we included the two most immediate contiguous scores as controls for each round number (e.g., we compared SAT scores of 1380 and 1390 jointly against 1410 and 1420 jointly, using 1400 as the baseline).
8. When we analyzed the scenarios separately, we obtained the same qualitative pattern of results in all three cases; the effect of distance from the round number was statistically significant for Scenarios 1 and 2 (ps < .001), but not for Scenario 3 (p = .4).

References

America's best colleges 1998. (1997, September). *U.S. News & World Report*, pp. 79–114.

Baseball Almanac. (n.d.). *Offensive stats 101*. Retrieved March 17, 2010, from http://www.baseball-almanac.com/stats.shtml

Feather, N.T. (1969). Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology*, *13*, 129–144.

Heath, C., Larrick, R.P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, *38*, 79–109.

Kahneman, D., & Miller, D.T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*, 136–153.

Lahman, S. (n.d.). *Sean Lahman's baseball archive*. Retrieved March 17, 2010, from http://www.baseball1.com/

Larrick, R.P., Heath, C., & Wu, G. (2009). Goal-induced risk taking in negotiation and decision making. *Social Cognition*, *27*, 342–364.

Medvec, V.H., Gilovich, T., & Madey, S.F. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology*, *69*, 603–610.

Medvec, V.H., & Savitsky, K. (1997). When doing better means feeling worse: The effects of categorical cutoff points on counterfactual thinking and satisfaction. *Journal of Personality and Social Psychology*, *72*, 1284–1296.

Mellers, B.A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision-affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science*, *8*, 423–429.

Pope, D.G., & Pope, J.C. (2009). The impact of college sports success on the quantity and quality of student applications. *Southern Economic Journal*, *75*, 750–780.

Retrosheet. (n.d.). *Play-by-play data files (event files)*. Retrieved March 17, 2010, from http://www.retrosheet.org/game.htm

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532–547.

Simonsohn, U. (2007). Clouds make nerds look good: Field evidence of the influence of incidental factors on decision making. *Journal of Behavioral Decision Making*, *20*, 143–152.

Vigdor, J.L., & Clotfelter, C.T. (2003). Retaking the SAT. *Journal of Human Resources*, *38*, 1–33.